Supplementary: Adiabatic Quantum Computing for Multi Object Tracking

Jan-Nico Zaech[1]   Alexander Liniger[1]   Martin Danelljan[1]   Dengxin Dai[1,2]   Luc Van Gool[1,3]

[1]Computer Vision Laboratory, ETH Zurich, Switzerland,
[2]MPI for Informatics, Saarbrucken, Germany, [3]KU Leuven, Belgium

{zaechj,alex.liniger,martin.denelljan,dai,vangool}@vision.ee.ethz.ch

1. Introduction

The supplementary material aims at giving a more thorough insight into the technical details of our work and at highlighting results obtained using simulated and quantum annealing. First, we prove that Hessian regularization does not influence the minimizer of the binary optimization problem in Section 2. After this, a further analysis of measurements generated using simulated and quantum annealing is presented in Section 4. Finally, detailed results on the MOT15 challenge are shown in Section 5 and furthermore, also visualized in the accompanying video.

2. Hessian Regularization

The following proof shows that the optimum solution is not influenced by the additional diagonal terms introduced in Section 5.1 of the main paper. This holds given a binary optimization problem and the constraints in Equations (6) and (7).

$$c_{ii} = \text{vct}(\mathbf{X}_i^T)\mathbf{E}_{ii}\text{vct}(\mathbf{X}_i) \qquad (1)$$

$$\overset{\text{diag}}{=} \sum_{t=1}^{T}\sum_{d=1}^{D} x_{id_it}^2 e_{d_it} \qquad (2)$$

$$\overset{\text{bin}}{=} \sum_{t=1}^{T}\sum_{d=1}^{D} x_{id_it} e_{d_it} \qquad (3)$$

$$= \sum_{t=1}^{T-1}\sum_{d=1}^{D} x_{id_it} e_{d_it} + \sum_{d=1}^{D} x_{id_iT} e_{d_iT} \qquad (4)$$

$$\overset{(22)}{=} \sum_{t=1}^{T-1}\sum_{d=1}^{D} x_{id_it} e + \sum_{d=1}^{D} x_{id_iT} 0 \qquad (5)$$

$$= e \sum_{t=1}^{T-1}\sum_{d=1}^{D} x_{id_it} \qquad (6)$$

$$\overset{(6)}{=} e \sum_{t=1}^{T-1} 1 = e(T-1) \qquad (7)$$

3. Post Processing

To allow the handling of long sequences that cannot be represented as a single optimization problem, the sequence needs to be split into overlapping subproblems. We split a long sequence in equally sized subproblems with an overlap similar to the modeled frame gap. After tracking each subproblem separately, tracks are matched between each pair of neighboring subproblems by solving a linear sum problem that can be solved in polynomial time. The optimization goal is to maximize the number of detections that are jointly assigned to tracks matched in both subproblems. The linear sum optimization problem for matching subproblems k and $k+1$ is stated as

$$\max_{x_{ij}\in\{0,1\}} \sum_{i=1}^{T_k}\sum_{j=1}^{T_{k+1}} x_{ij}\, m_{ij} \quad \text{s.t.} \quad \begin{array}{l} \sum_{i=1}^{T_k} x_{ij} \leq 1 \\ \sum_{j=1}^{T_{k+1}} x_{ij} \leq 1, \end{array} \qquad (8)$$

where x_{ij} are the optimization variables indicating an assignment of track i in segment k to track j in segment $k+1$, The considered tracks T_k and T_{k+1} are the tracks that have at least one detection assigned to them in the frames overlapping between both subproblems. m_{ij} is the number of detections shared by tracks i and j in the overlapping frames, which furthermore is set to a small negative value if tracks i and j have no overlap.

4. Annealing Energy Distribution

Results from simulated as well as quantum annealing with synthetic data are presented in Figures 1, 2, 3, and 4. In each of the figures, the topmost plot shows the probability of finding the correct solution and the plots below show the measurement energy for increasing noise levels from top to bottom.

Fixed Lagrangian. Results for a fixed Lagrangian multiplier are shown Figures 1 and 2 for real and synthetic data respectively. For simulated annealing, the Lagrangian multiplier is in the range $\lambda \in [2, 5]$ and with noise levels be-

	seq	MOTA	IDF1	MT	ML	FP	FN	IDs	Density [2]	Tracks [2]	Boxes [2]	FPS [2]
X-Val	Venice-2	41.6	50.0	13	1	2178	1855	135	11.9	26	7141	30
	KITTI-17	79.6	83.6	6	0	5	130	4	4.7	9	683	10
	KITTI-13	33.5	57.8	13	11	197	293	17	2.2	42	762	10
	ADL-Rundle-8	26.7	51.4	18	3	3587	1336	49	10.4	28	6783	30
	ADL-Rundle-6	63.3	53.7	11	1	228	1570	40	9.5	24	5009	30
	ETH-Pedcross2	46.2	59.9	28	74	127	3216	27	7.5	133	6263	14
	ETH-Sunnyday	78.1	87.0	19	6	110	295	2	5.2	30	1858	14
	ETH-Bahnhof	47.3	67.5	98	38	1933	895	24	5.4	171	5415	14
	PETS09-S2L1	83.2	76.9	17	0	341	351	58	5.6	19	4476	7
	TUD-Campus	75.5	75.4	4	0	9	72	7	5.1	8	359	25
	TUD-Stadtmitte	81.6	80.8	7	0	5	201	7	6.	10	1156	25
	OVERALL	59.7	67.6	234	134	8720	10214	370	7.3	500	39905	-
Test	Venice-1	44.4	49.0	6	3	656	1839	42	10.1	17	4563	30
	KITTI-19	48.2	60.1	14	17	528	2191	49	5.0	62	5343	10
	KITTI-16	52.7	67.1	3	1	120	666	19	8.1	17	1701	10
	ADL-Rundle-3	50.0	47.4	10	7	653	4346	81	16.3	44	10166	30
	ADL-Rundle-1	38.2	49.9	12	2	2365	3313	73	18.6	32	9306	30
	AVG-TownCentre	52.7	57.0	58	35	363	2767	250	15.9	226	7148	2.5
	ETH-Crossing	62.3	75.1	7	8	38	335	5	4.6	26	1003	14
	ETH-Linthescher	56.5	62.3	45	89	342	3493	48	7.5	197	8930	14
	ETH-Jelmoli	51.0	65.5	18	13	522	701	19	5.8	45	2537	14
	PETS09-S2L2	50.1	38.7	2	4	312	4259	243	22.1	42	9641	7
	TUD-Crossing	85.7	81.6	12	0	25	122	11	5.5	13	1102	25
	OVERALL	49.9	53.5	187	179	5924	24032	840	10.6	721	61440	-

Table 1. Results on the MOT15 [2] training and test set. Results on the training set are generated using leave-one-out cross validation (X-Val).

tween $\sigma = 0.2$ to $\sigma = 1.0$. For quantum annealing the ranges are $\lambda \in [1, 5]$ and $\sigma \in [0.0, 0.3]$ respectively.

For quantum as well as for simulated annealing, the spectral gap decreases with increasing noise levels, and thus, also the corresponding solution probability. When comparing the two approaches with each other, it becomes apparent that quantum annealing often returns high energy solutions, which corresponds to a higher temperature of the currently available systems.

Optimized Lagrangian. Results for optimized Lagrangian multipliers are shown in Figures 3 and 4 for real and synthetic data respectively. Noise parameters are the same as for fixed Lagrangian multipliers and the Lagrangian offset λ_{off} is in the range $\lambda_{\text{off}} \in [0, 2]$ for both approaches.

Following the same rules as for fixed Lagrangian multipliers, the spectral gap and corresponding solution probability decreases with increasing noise level. Comparing Figure 3 to Figure 1 reveals that in simulation a considerable improvement can be achieved by using optimized Lagrangian multipliers. Also for quantum annealing an advantage can be achieved, nevertheless, it is smaller than in simulated annealing, which can be explained by the higher noise level that results in high energy solutions.

5. MOTChallenge 2015

Detailed results for our method on each sequence in the MOT15 [2] training and test set are provided in Table 1. While the results on both sets are competitive with current state-of-the-art methods [1], the performance on the training set with leave one out cross-validation is higher than on the test set.

The difference can be explained by the harder examples represented by it. While both splits contain a similar number of frames (5500 frames and 5783 frames respectively), the number of tracks, detected boxes and the corresponding density is approximately 45% higher in the test set and thus, also the complexity and size of the optimization problem. As our formulation is designed for AQC using an Ising model, the resulting optimization problem is a quadratic binary program and thus, hard to solve on classical hardware. This becomes apparent for two sequences in the test set, *AVG-TownCentre* and *PETS09-S2L2* with a high density of 15.9 and 22.1 and low frame rate of 2.5 fps and 7 fps respectively. The two sequences account for only 27.3% of the total detections, but for 58.7% of the ID switches. ID switches are a good measure for the tracker's performance in this case, as they are less influenced by the performance
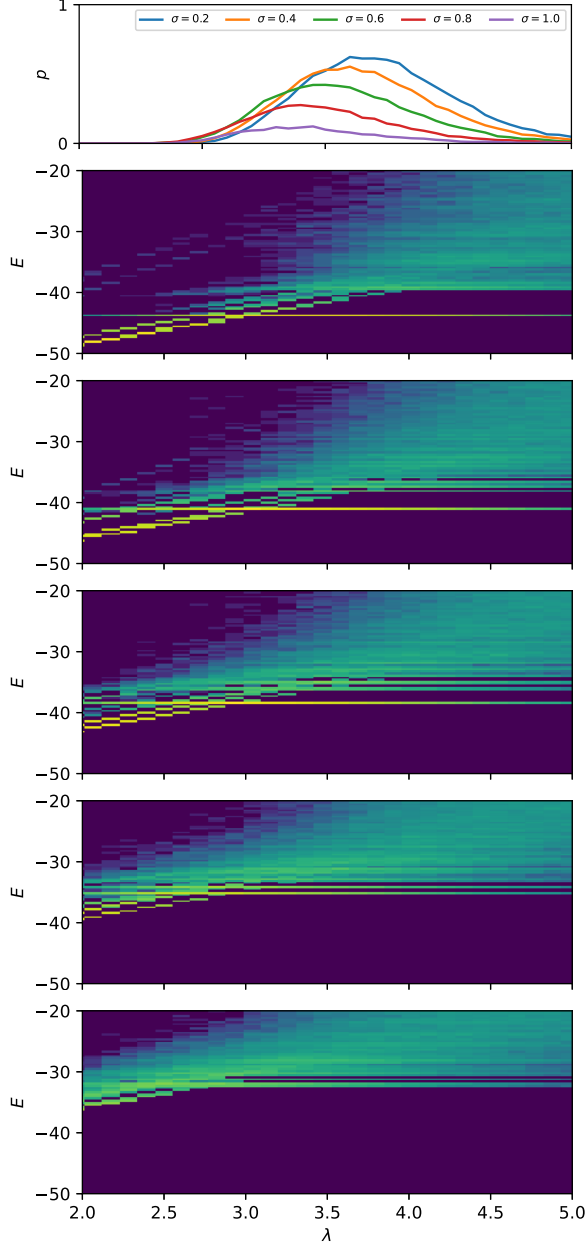
Figure 1. Solution probability and energy levels using simulated annealing for noise levels $\sigma \in \{0.2, 0.4, 0.6, 0.8, 1.0\}$ and changing λ.

Figure 2. Solution probability and energy levels using quantum annealing for noise levels $\sigma \in \{0.0, 0.1, 0.2\}$ over λ.

MOT15 [2], it can be expected that it scales up to larger datasets accordingly and thus, provides the basis to develop AQC based formulations of the MOT task.

References

[1] Andrea Hornakova, Timo Kaiser, Paul Swoboda, Michal Rolinek, Bodo Rosenhahn, and Roberto Henschel. Making Higher Order MOT Scalable: An Efficient Approximate Solver for Lifted Disjoint Paths. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, page 11, 2021. 2

[2] Laura Leal-Taixé, Anton Milan, Ian Reid, Stefan Roth, and Konrad Schindler. MOTChallenge 2015: Towards a Benchmark for Multi-Target Tracking. *arXiv:1504.01942 [cs]*, Apr. 2015. arXiv: 1504.01942. 2, 3

of the object detector than FP and FN. Due to the larger size of these problems, the optimization cannot finish for all segments within the given time frame and thus, returns a sub-optimal solution.

Even though the problem size is a limitation when solving the problem on classical hardware, it can be resolved when future AQCs become available. As the overall performance is similar to current state-of-the-art methods on
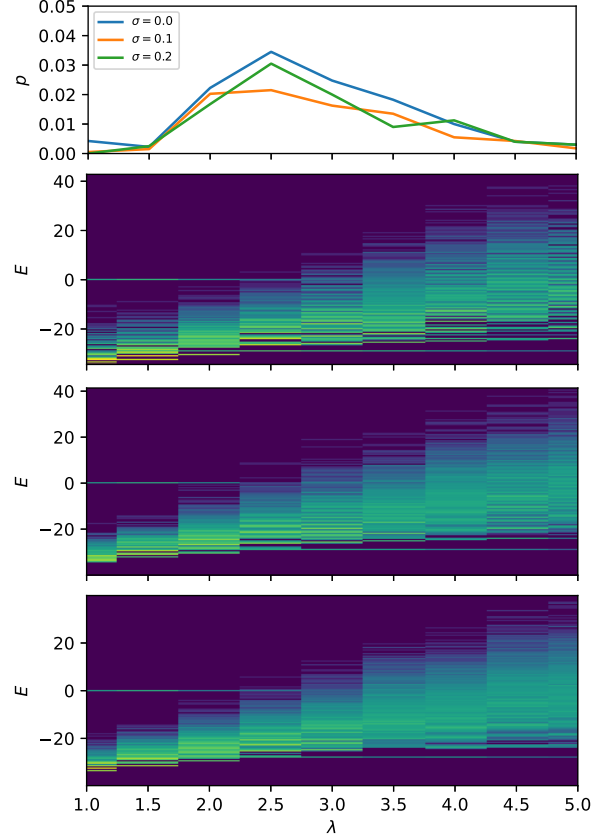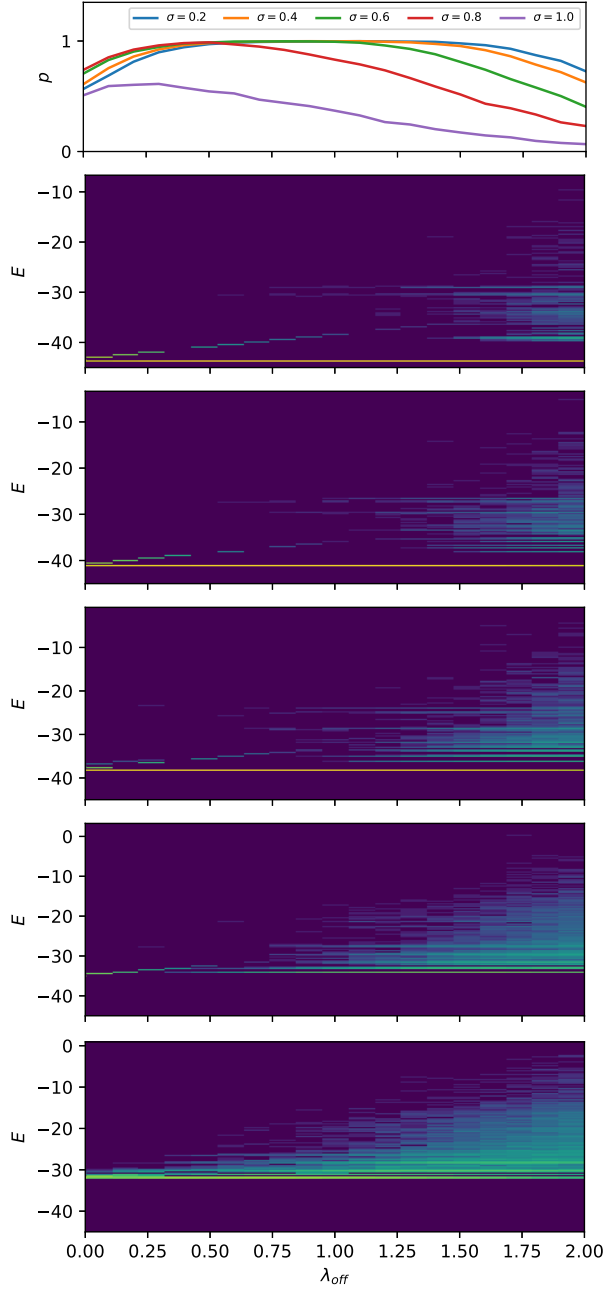
Figure 3. Solution probability and energy levels using simulated annealing and optimized λ_i for noise levels $\sigma \in \{0.2, 0.4, 0.6, 0.8, 1.0\}$ over λ_{off}.
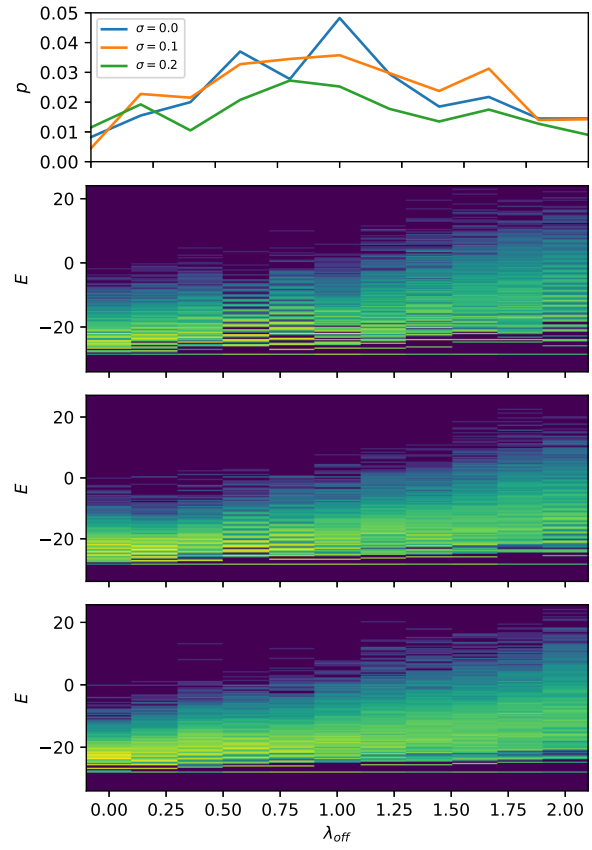

Figure 4. Solution probability and energy levels using quantum annealing and optimized λ_i for noise levels $\sigma \in \{0.0, 0.1, 0.2\}$ over λ_{off}.